全媒体传播中数据技术的应用实践

袁 然
[大众报业集团（大众日报社），山东 济南 250014]

摘 要：全媒体传播时代，数据正成为重要的生产要素。通过专注于数据产品及数据服务模式的创新，媒体行业不断实践着数据驱动的新变革。本文概述了大众报业集团在全媒体传播中数据采集、分析、呈现的技术实践，以及数据新闻、智库分析、数据评价、舆情服务等应用场景驱动的应用实践。思考了由此引发的数据安全、获取、真伪等方面的问题并提出了解决思路。

关键词：全媒体传播；数据要素；数据安全；数据获取；数据真伪　　**中图分类号：**G232　　**文献标识码：**A
文章编号：1671-0134（2021）07-021-03　　**DOI：**10.19483/j.cnki.11-4653/n.2021.07.004
本文著录格式：袁然．全媒体传播中数据技术的应用实践 [J]. 中国传媒科技，2021（07）：21-23.

习近平总书记指出：“全媒体不断发展，出现了全程媒体、全息媒体、全员媒体、全效媒体，信息无处不在、无所不及、无人不用，导致舆论生态、媒体格局、传播方式发生深刻变化，新闻舆论工作面临新的挑战。”[1]

全媒体时代，各领域产生了海量数据。数据所蕴含的价值、规律、趋势，正在重塑着传统媒体行业。特别是数据作为一种新型生产要素被人们广泛地重新认知，更加速了媒体深度融合发展的进程。

1. 数据正成为重要的生产要素

2020 年 4 月，《中共中央国务院关于构建更加完善的要素市场化配置体制机制的意见》，明确把数据作为五个生产要素之一，并构建以数据为关键要素的数字经济。党的十九届五中全会通过的《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》，提出建立数据资源产权，推动数据资源开发利用。推进数据等要素市场化改革。《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》，提出要“迎接数字时代，激活数据要素潜能”。在国家“十三五”规划中，大数据还是一个新兴的技术产业，在“十四五”规划中，大数据正在成为融入经济社会发展各领域的要素、资源、动力、观念。

2. 全媒体数据产业蓬勃发展

随着全媒体时代媒体深度融合发展的持续推进，全媒体数据产业得到蓬勃发展，吸引了越来越多的科研院所、民营资本的涌入。这些产业主体专注于大数据产品和服务，深耕媒体大数据基础技术研发和服务模式创新。

2.1 数据是洞察的眼睛

在如今“产品即服务”的时代，媒体数据服务商清博智能重视基于数据的服务产品。2020 年 2 月，清博推出疫情防护公益产品“疫情查”，在快速应对、打通政务、科研部门与民众的沟通方面都起到了积极作用。[2] 清博指数作为大数据评估平台，建立各垂直领域的影响力指数，注重传播能力考核。“新媒体管理考核系统”，贯通“策采编发评”流程，实现了新媒体矩阵一体化管理考核工作。清博通过大数据平台的全方位服务，用数据智能创新产业生态。

2.2 用数据智能推动社会进步

媒体数据服务商百分点科技注重新技术，构建了企业级、政府级和 SaaS 服务三大核心业务体系，形成了大数据、人工智能和在线服务三大类技术产品。提出基于大数据技术的全媒体中台建设思路。建设过程中需要考虑组织体系、管理体系、执行体系、技术体系。全媒体中台包括大数据中心资产平台、大数据中心能力平台、资源发布与展示平台、资源服务共享平台四部分。[3]

2.3 打造数据经济时代智能内容新生态

媒体数据服务如凡闻科技专注新闻内容大数据，打造智能内容新生态。提出了“大数据 + 人工智能 + 云服务”的创新模式。提供内容大数据技术服务、中央数据中心建设、全媒体生产流程改造、内容数据产品化变现和传播影响力分析等综合性服务。[4]

2.4 搜索信息、整合内容、发现知识

媒体数据服务商拓尔思 20 多年前便推出了全文检索和内容管理平台产品。进入全媒体时代，拓尔思全新打造了“以内容资产为核心”的智能生产与传播服务平台，重构新闻生产流程、提升数据价值挖掘与赋能，助力媒体融合的传播与服务变现。致力于以数据为媒体赋能，以知识为行业赋能。[5]

3. 全媒体传播数据处理实践

在全媒体传播实践中，数据存在于全生命周期，从数据采集、清洗、存储、挖掘、分析到可视化呈现，每一个环节都涉及大量的数据算法和数据处理方法。

3.1 全媒体传播中的数据采集打签

海量信息数据的产生给数据采集带来了挑战，数据采集的广度和精度决定了后续数据挖掘分析的深度。针对不同的需求，不同的数据类型都会有不同的采集方式和工具。通常数据有结构化、半结构化和非结构化之分，但在采集过程中，可以将数据进行更细粒度的划分并进行初步的数据标签化，以方便后期的数据挖掘和分析。

3.1.1 自主开发媒体数据采集工具

采用 web 请求机制，自主开发的网络爬虫软件，实

时获取互联网中的公开数据。这些数据大多是内容与用户的交互信息，是实时更新的海量数据，包含结构化和非结构化形式。通过设计关系型数据结构和非关系型数据结构进行数据存储。大众报业集团通过使用 python 语言自主开发软件爬虫工具，部署分布式服务器组，采集权威信源的内容数据，汇集到集团数据资源池，进行后续数据的开发利用。

3.1.2 数据接口方式采集数据

这种方式是通过数据接口直接与业务系统对接，获取业务系统的数据。这种方式需要多方协调，明确对方系统的业务流程以及接口规范等，前期对接、联调耗时大，但采集的数据精度高、实时性强。[3] 这种方式主要限于企业机构内部数据。大众报业集团通过与山东广电“闪电云”、商业平台等合作方数据接口对接，获取实时的原创稿件、热点新闻、热搜内容等数据。

3.2 全媒体传播数据分析呈现

通过对数据分析挖掘形成的趋势规律，是数据价值的体现。通过数据可视化手段将这些价值呈现出来也是数据实践的重要一环。大众报业集团持续推进数据在全媒体生产中的应用，丰富新闻作品内容形式。大众报业集团获中国新闻奖一等奖的作品《行走黄河滩・我的迁建故事》以数据素材为主轴，充分融合运用图解、VR、航拍等技术，形成融合创新的重大主题报道作品。该作品长卷页面使用的算法和新技术有以下几点：

3.2.1 长卷横向滚动展示法

通过设定页面高度，建立特定的显示区域，配置显示区域背景大图，设置显示区域的横轴为滚动方式，纵轴则隐藏超出显示区域的部分。在显示区域内再创建多个子显示区域，把子显示区域设置 inline-table 等属性参数。

3.2.2 新闻、视频、VR 和航拍素材链接在长卷上的定位方法

通过定义不同的数组记录不同的事件及图标，在数组内记录相对位置数值。计算位置时需要把背景大图的总宽度与当前屏幕宽度组合成新的坐标系，精确计算当前时间各事件、图标在展示画面中的绝对位置。

3.2.3 多端兼容展示方法

通过 CSS 样式代码控制兼容显示效果，将各图片按比例缩放适配手机屏幕的宽度和高度。计算相对位置时，以背景大图总宽度为坐标系，动态获取并计算具体位置数据，还要兼容手机端和 PC 端的不同尺寸。

4. 全媒体传播数据应用实践

作为新闻信息内容生产者的现代传媒集团，大众报业集团持续在全媒体传播数据方面不断进行尝试和探索。

4.1 数据新闻

数据新闻已经成为媒体日常的标配，其着重点有“数据分析”“内容策划”和“可视呈现”等，根本还是在讲好新闻的故事，数据作为故事的基础，要有结构化和知识化的梳理。大众报业集团积极探索具有交互式、可视化的数据新闻，融入重大主题报道策划，持续推出“辉煌十三五”“牢记嘱托 走在前列 全面开创”“疫情防控”“2020 全国两会”“2020 山东省防汛应急指挥大数据”等系列数据新闻产品。在融媒体产品生产中形成技术与内容融合的独特生产方式。[6]

4.2 智库分析服务

通过采用大数据分析技术，配合资深内容采编人员进行深度分析，大众报业集团每月推出《中央媒体对山东报道选编》的数据智库分析报告，报送山东省委宣传部及省主要领导。通过自动监测中央媒体对山东的报道，赋权运算，自动抓取关注度高的文章和视频并整理入库，再经过平台算法匹配和分析，从中央媒体报道角度切入，形成包含重磅聚焦点、重点稿件传播数量和路径、网民热度、情感倾向、分布区域等内容的大数据智库分析报告。以此为基础，通过平台定义的多种主题模板，迅捷生成各种不同的数据智库产品。

4.3 数据评价完善考核机制

全媒体传播中的绩效考核更具复杂性也更体现人性化。大众报业集团的编辑记者考核条例已实行多年，在媒体融合机制下，主力军要进入主阵地，因此新技术和新媒体的运用成为新方向。为体现指挥棒引导作用，引入互联网评价数据，根据原创稿件数量、质量、传播效果、影响力等多维度指标，统计采编、运营人员及其部门等的工作量和工作效果，依据事先设立的考核公式计算生成稿费统计清单。使绩效指标清晰量化，考核结果公正公开，为量化管理提供大数据支持。

4.4 数据舆情服务

全媒体传播中，舆情的形成和发展都能在舆情服务系统的监测下一目了然。大众报业集团一直重视这方面的研发，经过多年的不断迭代和更新，目前的舆情系统服务于集团各媒体，同时舆情服务输出后已覆盖山东主要厅局。通过舆情大数据平台，可以清晰快捷地了解全网实时热点、地域分布和传播路径。通过不间断地监测，可以对文字、链接、图片、视频、音频等多种数据类型进行采集，从海量信息中获取舆情信息，过滤掉大量无效数据，再通过人工研判，及时预警并定向跟踪。

5. 全媒体传播数据实践思考

5.1 数据安全的思考

2021 年 6 月 10 日，《数据安全法》经十三届全国人大常委会第二十九次会议通过，并将于 2021 年 9 月 1 日起施行。这是数据领域的基础性法律，也是国家安全领域的一部重要法律。大数据、云计算的发展更依赖于开放和共享，与此同时也面临安全的困扰，这不仅仅是数据技术的问题，更是安全意识的问题。媒体的未来，是数据能力的较量，数据安全是数据能力的基础，除了云、网、平台、数据等基础方面的技术安全，还需要考虑数据安全本身的敏感性、隐私、涉密、版权等方面的法律安全，需要综合运用法律、制度、标准、技术等形成数据安全防护体系。

5.2 数据获取的思考

从网络公开的信息中获取有价值的数据是件很难的事，一是数据格式不规范，二是数据开放程度不一。格式的问题，一方面需要对抓取来的数据进行处理、清洗过滤、统一格式，另一方面可以设计推行数据标准逐步规范网络数据。针对数据开放的问题，党的十九届五中全会通过的《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》提出将扩大基础公共信息数据有序开放，建设国家数据统一共享开放平台。由党和国家在顶层推动，相信数据开放度会越来越高。对有数据产权的信息，通过购买或者合作，可以更高效获取需要的数据。

5.3 数据真伪的思考

全媒体时代，人人都是麦克风，随着大量自媒体发布的大量数据充斥网络，导致网络中的数据可信度不高，虽然权威信源能大幅提高信息数据准确性，但数据时效性又大打折扣，而且不能保证完全正确、准确。通过多信源对比来提高数据的准确性是可行的方法。如果通过引入区块链技术建立可信数据互信机制，经过认证的数据是可信的，这将大大减少人们获取正确数据的成本。

结语

全媒体传播中数据带来的新技术、新模式、新业态，正逐渐推动媒体深度融合创新发展。在相关政策的扶持和引导下，越来越多的媒体进军全媒体传播数据领域，越来越多的科研院所投入研发力量，孵化出数据服务模式和实体公司。虽然在技术和应用上还存在着成长的烦恼，但全媒体传播随着数据实践的不断深入，越来越多的新应用、新产品、新服务被开发出来，成为媒体深度融合发展的新动力。媒

参考文献

[1] 习近平 . 加快推动媒体融合发展 构建全媒体传播格局 [J]. 求是，2019（6）：4-8.

[2] 苗志如 . 浅析重大公共卫生事件中大数据的应用开发要点——以清博大数据为例 [J]. 中国传媒科技，2020（3）：19-21.

[3] 杨东，滕健 . 基于大数据技术的全媒体中台实践 [J]. 中国传媒科技，2020（9）：19-20.

[4] 王鸣 . 大数据 + 人工智能 + 云服务：技术创新驱动县级融媒体中心建设 [J]. 传媒，2019（2）：17-18.

[5] 林松涛 .AI 赋能媒体的场景创新 [J]. 中国传媒科技，2020（4）：10-12.

[6] 汤代禄 .5G 融媒技术探索与机制思考 [J]. 中国传媒科技，2020（8）：7-10.

作者简介：袁然（1972-），山东寿光，高级工程师，大众报业集团（大众日报社）信息技术部。

（责任编辑：陈旭管）

